

Fair draws for group rounds in sport tournaments

Pablo Laliena^a and F. Javier López^b 

^a*Dpto. Métodos Estadísticos, Facultad de Ciencias, Universidad de Zaragoza.
C/ Pedro Cerbuna, 12, 50009 Zaragoza, Spain*

^b*Dpto. Métodos Estadísticos and BIFI, Facultad de Ciencias, Universidad de Zaragoza.
C/ Pedro Cerbuna, 12, 50009 Zaragoza, Spain
E-mail: 648478@unizar.es [Laliena]; javier.lopez@unizar.es [López]*

Received 9 August 2017; received in revised form 17 January 2018; accepted 12 May 2018

Abstract

We propose two draw systems for the group round of sport tournaments where there are some geographical and/or seeding restrictions. One of the systems, related to the equal-sum partition problem, is “perfect,” since it yields perfectly balanced groups. The other system, which uses the classical scheme of extracting teams from pots, is heuristic and gives results where the groups have very similar scores. We apply our results to Fédération Internationale de Football Association (FIFA) Soccer World Cups and show that our proposals are much better than the FIFA system and also outperform other recently developed systems.

Keywords: sports; scheduling; heuristics; assignment problems; integer programming

1. Introduction

It is well known that mathematics and sports are no strangers to one another. In particular, operational research has been widely used in sports scheduling; see, for instance, Ribeiro (2012), Alarcón et al. (2014), or Durán et al. (2017) for some recent contributions in this area. In the present paper, we will focus on the draw systems used for the group rounds in sport tournaments.

Many tournaments in sports, such as soccer or basketball, are organized with a group round followed by a knockout phase. Examples are FIFA Soccer World Cup, UEFA Soccer Champions League, FIBA Basketball World Cup, Rugby World Cup, Cricket World Cup, among many others. In the group round, teams are assigned to groups and, after playing against all others in the group, the best performing teams in each group qualify for the next round. The allocation of teams in groups is subject to some rules, such as the existence of seeds (highest ranked teams in tournament are assigned to different groups) or geographical rules (e.g., no two teams from the same country can be assigned to the same group in UEFA Soccer Champions League).

The procedure for creating the groups is usually a draw system designed to satisfy the established rules. On most occasions, the system involves drawing balls bearing the names of teams from urns, in order to provide a sense of excitement and anticipation and to prove transparency. However, in many tournaments, the draw systems yield results that are considered poor, since the resulting groups are very different in terms of strength. This implies that some strong teams, allocated in a harsh group will have more difficulties in qualifying than other weaker teams, allocated in an easier group, which is regarded as unfair. The main problem when designing a draw system is to guarantee that it yields balanced groups while satisfying the seeding, geographical, or other rules imposed by the organizer.

One example, perhaps the most archetypal, of a tournament where this problem exists is the FIFA Soccer World Cup. This tournament is played by 32 teams every four years. The tournament begins with a group round, with eight groups of four teams each. Two teams from each group qualify for the next round. FIFA uses a draw system for placing the teams in groups: teams are introduced into pots and they are randomly extracted and assigned to groups following certain rules. The composition of the pots depends on the geographical areas of the teams as well as on their FIFA ranking.

The FIFA draw system has been widely criticized for several reasons. One of these reasons is the FIFA ranking system, which has been shown to have less predictive power than other ranking systems; see Lasek et al. (2013) and McHale and Davies (2007). Beyond the ranking system, many criticisms are for the draw system; see, for example, Jones (1990) for an analysis of the flaws in the system used by FIFA to create the groups in the 1990 World Cup. Criticisms reached a new peak in the 2014 World Cup; see, for instance, McMahan (2013). One important flaw of the FIFA draw system is that some teams are “penalized”: despite having better rankings than other teams they are more likely to end up in a stronger group, see Aisch and Leonhardt (2014). Another important reason to seek balanced groups is that weak groups will be less profitable in terms of audience. Moreover, very strong groups, having more than two top teams, will prevent at least one of them from qualifying to the knockout stages, resulting in a less interesting tournament. All of these flaws and criticisms are gathered in Guyon (2015), where the author analyzes in detail the problems of the draw system. He also proposes other types of systems that would solve these flaws, and yield a more equitable result. All these flaws and criticisms are gathered in Guyon (2015), where the author analyzes in detail the problems of the draw system. He also proposes other types of systems that would solve these flaws, and yield a more equitable result.

Since the FIFA Soccer World Cup is probably the most important sport tournament in the world (more than one billion people watched the 2014 final game, according to Kantar Media, 2015), we will focus on it. In particular, we consider the problem of allocating 32 teams in 8 groups where two teams from each group qualify for the next round, subject to the existence of seeds and some geographical rules. Our systems can also be used, with some modifications, for a different number of teams or groups, or different rules. The seeding system used by FIFA implies that the host team and the seven strongest teams in the tournament are assigned to different groups. In order to guarantee that groups are formed by teams from different geographical areas, FIFA establishes the so-called geographical rule. It defines six qualifying zones: Africa, Asia, Europe, North/Central America and Caribbean, Oceania, and South America (we will call them “continents” for simplicity). The geographical rule states that team from the same continent (except for Europe) cannot be in the same group; also, no more than two European teams can be in the same group. To fulfill the geographical

rule, teams are placed in pots according to their continent, and balls are drawn from the pots. For a thorough explanation of the system, the reader can consult Guyon (2015). The main criticism of the FIFA draw system is that it usually produces results with groups having very different strengths. For instance, in the 2014 World Cup, group D, composed of Uruguay (6), Italy (9), England (10), and Costa Rica (31), was very strong, while group H, composed of Belgium (5), Russia (19), Algeria (32), and Korea (56), was much weaker (the numbers in parentheses are the FIFA rankings of the teams; the lower the number, the stronger the team). Another problem with the FIFA system is that it is not evenly distributed. That is, not all the possible outcomes of the draw have the same probability, which is considered unfair (see Section 2 in Guyon, 2015).

One important aspect to be considered is the definition of the strength of a group. We assume that we have a ranking of the teams such as FIFA ranking, ELO ranking, or others. One way to measure the strength of a group is to add the rankings of the four teams in it: the lower the sum, the stronger the group. However, since only two teams qualify for the next round, a group is usually considered tough when three teams are strong, even if the fourth one is much weaker. In fact, Guyon (2015) argues that the sum of the ranks of the three strongest teams in a group is considered as a more relevant measure of the strength of the group than the sum of the ranks of its four teams. Therefore, in what follows, we will define the strength of a group as the sum of the rankings of its three strongest teams. See Remark 1 for further explanation on this definition.

Guyon (2015) proposes three different draw systems in order to get groups of similar strength satisfying the seeding and geographical rules for FIFA World Cups. His proposals had a big impact in the media, see Aisch and Leonhardt (2014) and Guyon (2014). Among these systems, he selects the so-called S-curve-type system. He shows that his method is evenly distributed, easy to implement, and much better than the FIFA draw system. Note, however, that Guyon's method does not always produce groups with exactly the same strength.

Our main objective is to develop a draw system that is evenly distributed and where all the groups have the same strength while satisfying the FIFA geographical rule and seeding system. Of course, a way to achieve this would be to list all the possible results (a result is the distribution of the 32 teams in 8 groups) and to analyze, for each one, if all the groups have the same strength and satisfy the geographical rule. Once all the results have been listed, a uniform draw among them would be used to define the groups for the World Cup. As the seeds are already fixed in groups A–H, there are $24!/(3!)^8 \sim 3.69 \times 10^{17}$ possible results to be checked, which is too high for real-time computation. Guyon, after a careful analysis, shows that, in the particular case of the 2014 World Cup, this number can be reduced to 7.84×10^{11} , but this is due to the particular number of teams from each continent and to the fact that four South-American and four European teams are the seeds. For future World Cups, this may change and different numbers will most likely appear. Moreover, in order to apply that reduction, a thorough analysis as the one carried out in Guyon (2015) would be necessary for each tournament. The method we propose is “universal,” in the sense that it does not depend on the number of teams from each continent nor on the continent of the seeds. Our method will first list all the results with even groups (all the groups will have exactly the same score) and then it will check the geographical constraints in each of them. Once all the feasible solutions (results with even groups satisfying the geographical rule) have been listed, one of them is picked at random via a uniform draw, thus guaranteeing the equiprobability of the results.

One aspect of our method, which may not be very appealing, is its practical implementation. The number of feasible solutions may be very large, so the final draw, on live TV, would consist of

drawing a ball from an urn with a very large number of balls. For the 2014 World Cup, using relative ranks 1, . . . , 32 (based on FIFA ranking), the number of feasible solutions is 173,160. Of course, such a large number of balls is unreasonable for the TV show. A possibility would be to apply the well-known rejection method: balls from 0 to 9 are drawn for units, tens, hundreds, thousands, and ten thousands, while a ball from 0 to 1 is drawn for hundred thousands; if the result is between 1 and 173,160 then it is accepted, otherwise it is rejected and all the extractions are repeated until a valid number is obtained. The list with all the solutions should be published in advance on the FIFA web page. Another possibility would be to reduce this number before the final draw using some objective criteria; for instance, as our method produces results where the groups have the same score defined as the sum of the three strongest teams, keep only those where the sum of the four teams is very similar in all groups. Another possibility is the following: list all the feasible solutions on the FIFA web page and, some weeks before the final draw, pick a small number of them at random. Those will be the only solutions for the final draw on the TV show. As these solutions would be available before the show, they would attract much attention in the media.

In case a draw system based on extracting teams from pots is preferred, we also define an alternative system which may be more appealing. This second system, called the 3-2-3 draw system, is heuristic and does not guarantee that all the groups have the same strength, but it still produces very good results. It is based on the S-curve-type algorithm in Guyon (2015) and can be seen as an improvement of that system.

Remark 1. *Measuring the strength of a group via its three strongest teams.* While there is a degree of subjectivity in the definition of strength of a group, we have decided to take it as the sum of the rankings of its three strongest teams, as suggested by Guyon (2015), instead of the sum of its four teams. We believe that this measure is more relevant when trying to get balanced groups. Of course, a group with teams ranked {1, 9, 10, 11} is stronger than a group with rankings {1, 9, 10, 32}. However, since there are weak teams in the tournament, the search for balance will imply that weak teams are placed in different groups, so the former group will not be a possible result of the draw. For instance, both in Guyon's and in our algorithms, it is guaranteed that every group has at least one team among the 12 weakest ones. Also, current draw systems such as FIFA's usually yield groups having a weak team; in fact, none of the weakest teams in their groups reached semifinals in any of the FIFA World Cups in the period 1994–2014 (FIFA rankings were created in 1993). The use of the three strongest teams in the definition will avoid that the value of the strength of a group is driven by its weakest team. Indeed, using the four teams in the definition may lead to some undesired situations; for instance, if we take absolute rankings, a group {3, 9, 10, 59}, $strength = 81$, having 3 top 10 teams, will be considered as weaker than a group {3, 9, 32, 34}, $strength = 78$ (both groups satisfy 2014 World Cup seeding and geographical rules). This effect is somewhat mitigated if we use relative rather than absolute rankings, but in this situation, group {4, 9, 10, 32}, $strength = 55$ (having 3 top 10 teams), will be considered as weaker than a group {4, 9, 19, 22}, $strength = 54$. This problem is even more clearly seen when groups are formed by more teams, such as in FIFA World Cup Qualifiers or tournaments in other sports. For instance, in 2019 Rugby World Cup, groups will be formed by five teams and two of them will qualify for the next round: group C, having England (2), France (6), and Argentina (9), was considered as a very strong group even before knowing the other teams in it, see Menezes (2017). Finally, note that in popular culture, a “group of death,” considered as a very strong group, can be defined as “a group which is unusually

competitive, because the number of strong competitors in the group is greater than the number of qualifying places available for the next phase of the tournament” (Wikipedia, 2017). This was the case with group H in the 2017–2018 UEFA Champions League, formed by Real Madrid, Tottenham Hotspur, Borussia Dortmund, and Apoel, unanimously declared as the “group of death” by the strength of three of its teams; see, for example, Law and Bagchi (2014).

We remark that, while our draw systems are designed to achieve the best results in terms of balancedness with the definition of strength as the sum of the three best teams, they also perform very well if we analyze the results with the metric of the four teams in the group (see Section 4).

Remark 2. *Absolute rankings versus relative rankings.* When defining the score of a group, we add the ranking of the three strongest teams. We can use either absolute or relative rankings. Guyon (2015) argues that the use of absolute FIFA rankings in the definition of strength can be misleading since the ranks can vary from 1 to more than 200. Therefore, he works with relative rankings from 1 to 32. We develop our “perfect” system in Section 2 in a more general setting, where the ranks of teams are arbitrary positive integer numbers. We will analyze, as a particular case, the situation where the ranks are $1, \dots, 32$. For our second system in Section 3, we use only the relative ranks $1, \dots, 32$, in order to easily compare it with Guyon’s proposal.

The paper is organized as follows. The perfect draw system and the 3-2-3 draw system are described in Sections 2 and 3, respectively. Section 4 is devoted to the comparison of our draw systems with other existing procedures. The conclusions of the work are presented in Section 5.

2. The perfect draw

We now explain our first algorithm that will produce groups with exactly the same score. As described in the Introduction, the score of a group will be the sum of the rankings of its three strongest teams. Let E , a set of 32 positive integer numbers, be the set of rankings of the teams. An important particular case, to be analyzed in Section 2.2, is $E = \{1, \dots, 32\}$, that is, when we use the relative ranks. A result is a partition of E into eight subsets with four elements each. By the FIFA seeding rule, the eight smallest elements of E must be in different subsets. We begin with some definitions.

Definition 1. *A grouping $\mathbf{x} = (x_1, \dots, x_8)$ is an 8-tuple of elements in E^4 , with $x_j = (l_j, m_j, n_j, p_j)$, for $j = 1, \dots, 8$, such that*

- (i) $l_1 < \dots < l_8$ are the 8 smallest elements of E ;
- (ii) $m_j < n_j < p_j$, for $j = 1, \dots, 8$.

Each element x_j of \mathbf{x} is a group and the score of group x_j , for $j \in \{1, \dots, 8\}$, is defined as $s_j = l_j + m_j + n_j$.

Note that there is a bijection between the set of results of placing 32 teams in 8 groups (with a seed in each group) and the set of groupings.

Definition 2. *A grouping \mathbf{x} is a perfect solution if $s_1 = \dots = s_8$. In this case, we say that the value of the solution is s_1 .*

If we find a perfect solution, we ensure that all groups have the same score. As we want the strong teams to be in different groups, we look for perfect solutions whose value is minimum, which would be achieved if the eight weakest teams were in different groups. Thus, the minimum value of a perfect solution will be greater than or equal to the sum of the 24 smallest elements of E divided by 8. In the case of $E = \{1, \dots, 32\}$, this quantity is $(24 \times 25/2)/8 = 37.5$, so the lowest value of a perfect solution in this case will be, at least, 38.

The problem above can be written as follows: “Let E be a set of 32 positive integers and s a positive integer: find all the partitions of E into 8 subsets of size 4 such that the 8 smallest elements of E are in different subsets and the sum of the three smallest elements of each subset is equal to s .” This problem is related to the equal-sum partition problem, which looks for partitions of a set into subsets of equal sum. That problem has been studied in detail in the literature; see, for instance, Korf (2009), Michiels et al. (2003), Prodinger (1982, 1984), Woeginger and Yu (1992), Zhang et al. (2011), and it is NP-hard. Note, however, that there are two important differences between our problem and the equal-sum partition problem. First, we have one fixed element in each subset. Second, and more important, it is not the sum of the four elements of the subsets but the sum of their three smallest elements which must be equal.

This problem can be easily formulated as a binary linear programming problem, with 448 variables and 772 constraints. In the particular case $E = \{1, \dots, 32\}$, it took more than two minutes to find one perfect solution with value 38 when solved by Gurobi optimization software on an AMD Athlon 64 3800+ 2.4 GHz. However, we want all the solutions of the problem; for this particular case, there are 7,629,120 solutions (173,160 of them satisfy the geographical rule), so we cannot expect to obtain them in a reasonable time using the binary programming formulation. We therefore provide an algorithm for finding all the perfect solutions.

In order to describe our algorithm, we need two more definitions.

Definition 3. *A configuration is a partition of E into four sets (B_1, B_2, B_3, B_4) of eight elements such that*

- (i) B_1 is formed by the eight smallest elements in E ,
- (ii) $\min\{m \in B_2\} < \min\{n \in B_3\}$,
- (iii) $\max\{m \in B_2\} < \max\{n \in B_3\}$.

Definition 4. *We say that a grouping \mathbf{x} is generated by a configuration (B_1, B_2, B_3, B_4) if $l_j \in B_1$, $m_j \in B_2$, $n_j \in B_3$ and $p_j \in B_4$, for $j = 1, \dots, 8$.*

Note that every grouping is generated by a unique configuration, namely $B_1 = \{l_1, \dots, l_8\}$, $B_2 = \{m_1, \dots, m_8\}$, $B_3 = \{n_1, \dots, n_8\}$, and $B_4 = \{p_1, \dots, p_8\}$. The fact that we require condition (ii) in Definition 1 guarantees that no other configuration can generate the same grouping. On the other hand, a configuration may not generate any grouping, again by condition (ii) in Definition 1. Given a configuration, it is easy to compute all the groupings that it generates. Each group is formed by an element from each set B_i , as long as the element from set B_i is smaller than the element from set B_{i+1} , $i = 2, 3$.

Lemma 1. *Let v be the sum of the elements of E . If (B_1, B_2, B_3, B_4) generates a perfect solution with value s , then the sum of the elements of B_4 is $v - 8s$.*

Proof. It is straightforward by noting that the score of any grouping generated by (B_1, B_2, B_3, B_4) is the sum of its elements in $B_1, B_2,$ and B_3 . \square

The above lemma specifies which sets B_4 are to be considered for the generation of perfect solutions with value s . It is a simple problem, which can be solved by dynamic programming or direct enumeration, to find, from 32 integers, all the subsets of eight elements whose sum is a given quantity. Therefore, in order to find all the perfect solutions with value s , we first find all the sets B_4 and, for each B_4 , find all the B_2, B_3 which generate perfect solutions.

Definition 5. A configuration (B_1, B_2, B_3, B_4) is invalid for value s if it verifies any of the following conditions:

- (i) $\max\{l \in B_1\} + \max\{j \in B_2 \text{ s.t. } j < \min\{k \in B_3\}\} < s - \min\{k \in B_3\},$
- (ii) $\min\{l \in B_1\} + \min\{k \in B_3 \text{ s.t. } k > \max\{j \in B_2\}\} > s - \max\{j \in B_2\}.$

Proposition 1. If a configuration (B_1, B_2, B_3, B_4) is invalid for value s , it does not generate any perfect solution with value s .

Proof. Suppose (i) is satisfied and let (x_1, \dots, x_8) be a grouping generated by (B_1, B_2, B_3, B_4) . Let x_i be the group where $\min\{k \in B_3\}$ is. Since the score of any group is the sum of the elements of $B_1, B_2,$ and B_3 , the maximum possible score of x_i is $\max\{l \in B_1\} + \max\{j \in B_2 \text{ s.t. } j < \min\{k \in B_3\}\} + \min\{k \in B_3\} < s$. Therefore, (x_1, \dots, x_8) is not a perfect solution with value s . An analogous argument works for (ii). \square

For convenience we say that a configuration is *valid* for value s if it is not invalid for value s in the sense of Definition 5. Note however that a valid configuration for value s may not generate any perfect solution with value s . Also, as the value s will be fixed, we will omit “for value s ” in what follows when saying that a configuration is valid or invalid.

Definition 6. Let $B_4 \subset E \setminus B_1$. The basic configuration associated to B_4 is (B_1, B_2^0, B_3^0, B_4) where B_2^0, B_3^0 are such that $\max\{j \in B_2^0\} < \min\{k \in B_3^0\}$.

That is, fixed B_4 , in the basic configuration B_2^0 is formed by the eight smallest elements of $E \setminus (B_1 \cup B_4)$ and B_3^0 by the rest.

Definition 7. For a basic configuration (B_1, B_2^0, B_3^0, B_4) a change of r teams is $(\{j_1, \dots, j_r\}, \{k_1, \dots, k_r\})$, where $\{j_1, \dots, j_r\} \subset B_2^0$ with $j_1, \dots, j_r \neq \min\{j \in B_2^0\}$ and $\{k_1, \dots, k_r\} \subset B_3^0$ with $k_1, \dots, k_r \neq \max\{k \in B_3^0\}$. We say that a configuration (B_1, B_2', B_3', B_4) is the result of the change $(\{j_1, \dots, j_r\}, \{k_1, \dots, k_r\})$ if $B_2' = B_2^0 \setminus \{j_1, \dots, j_r\} \cup \{k_1, \dots, k_r\}$ and $B_3' = B_3^0 \setminus \{k_1, \dots, k_r\} \cup \{j_1, \dots, j_r\}$. We say that a change of r teams is invalid for value s if the resulting configuration is invalid (for value s).

In the definition above we have required $j_1, \dots, j_r \neq \min\{j \in B_2^0\}$ and $k_1, \dots, k_r \neq \max\{k \in B_3^0\}$ as otherwise the result would not be a configuration due to conditions (ii) and (iii) of Definition 3. Note that, given B_4 , any configuration (B_1, B_2, B_3, B_4) is the result of a change of r teams ($r = 0, 1, \dots, 7$) from the basic configuration (B_1, B_2^0, B_3^0, B_4) . As above, we say that a change is *valid*

for value s when the resulting configuration is valid (for value s). Again, we will omit “for value s ” when saying that a change is valid or invalid.

Definition 8. Let (B_1, B_2^0, B_3^0, B_4) be a basic configuration and let $(\{j_1\}, \{k_1\})$ and $(\{m_1\}, \{n_1\})$ be two changes of one team. We say that $(\{m_1\}, \{n_1\})$ is stronger than $(\{j_1\}, \{k_1\})$ if any of the two following conditions is satisfied:

- (i) $j_1 = m_1$ and $k_1 < n_1$, or
- (ii) $k_1 = n_1$ and $j_1 > m_1$.

The following proposition is the basis of our algorithm.

Proposition 2. Let (B_1, B_2^0, B_3^0, B_4) be a basic configuration. We have the following.

- (a) If the change $(\{j_1\}, \{k_1\})$ is invalid and $(\{m_1\}, \{n_1\})$ is stronger than $(\{j_1\}, \{k_1\})$ then $(\{m_1\}, \{n_1\})$ is invalid.
- (b) Let $t < r$, $\{j_1, \dots, j_t\} \subset \{j_1, \dots, j_r\}$ and $\{k_1, \dots, k_t\} \subset \{k_1, \dots, k_r\}$. If the change $(\{j_1, \dots, j_t\}, \{k_1, \dots, k_t\})$ is invalid then the change $(\{j_1, \dots, j_r\}, \{k_1, \dots, k_r\})$ is invalid.
- (c) If the configuration (B_1, B_2^0, B_3^0, B_4) is invalid then any change is invalid.

Proof. (a) Let $(\{j_1\}, \{k_1\})$ be an invalid change and $(\{m_1\}, \{n_1\})$ a stronger change. Let (B_1, B_2, B_3, B_4) be the result of the change $(\{j_1\}, \{k_1\})$ and (B_1, B_2', B_3', B_4) the result of the change $(\{m_1\}, \{n_1\})$. Suppose that $j_1 = m_1$, so $k_1 < n_1$. We have

- $\min\{k \in B_3\} = \min\{k \in B_3'\} = j_1$;
- $\max\{j \in B_2 \text{ s.t. } j < \min\{k \in B_3\}\} = \max\{j \in B_2^0 \text{ s.t. } j < j_1\} = \max\{j \in B_2' \text{ s.t. } j < \min\{k \in B_3'\}\}$;
- $\max\{j \in B_2\} = k_1 < \max\{j \in B_2'\} = n_1$;
- $\min\{k \in B_3 \text{ s.t. } k > \max\{j \in B_2\}\} = \min\{k \in B_3^0 \text{ s.t. } k > k_1\} < \min\{k \in B_3^0 \text{ s.t. } k > n_1\} = \min\{k \in B_3' \text{ s.t. } k > \max\{j \in B_2'\}\}$.

Therefore, if (i) or (ii) in Definition 5 is satisfied for (B_1, B_2, B_3, B_4) so it is for (B_1, B_2', B_3', B_4) . The proof for the case $j_1 > m_1, k_1 = n_1$ is analogous.

(b) Let (B_1, B_2, B_3, B_4) be the result of the change $(\{j_1, \dots, j_t\}, \{k_1, \dots, k_t\})$ and (B_1, B_2', B_3', B_4) be the result of the change $(\{j_1, \dots, j_r\}, \{k_1, \dots, k_r\})$. We have

- $\min\{k \in B_3\} = \min\{j_1, \dots, j_t\} \geq \min\{j_1, \dots, j_r\} = \min\{k \in B_3'\}$;
- $\max\{j \in B_2 \text{ s.t. } j < \min\{k \in B_3\}\} = \max\{j \in B_2^0 \text{ s.t. } j < \min\{j_1, \dots, j_t\}\} \geq \max\{j \in B_2^0 \text{ s.t. } j < \min\{j_1, \dots, j_r\}\} = \max\{j \in B_2' \text{ s.t. } j < \min\{k \in B_3'\}\}$;
- $\max\{j \in B_2\} = \max\{k_1, \dots, k_t\} \leq \max\{k_1, \dots, k_r\} = \max\{j \in B_2'\}$;
- $\min\{k \in B_3 \text{ s.t. } k > \max\{j \in B_2\}\} = \min\{k \in B_3^0 \text{ s.t. } k > \max\{k_1, \dots, j_t\}\} \leq \min\{k \in B_3^0 \text{ s.t. } k > \max\{k_1, \dots, k_r\}\} = \min\{k \in B_3' \text{ s.t. } k > \max\{j \in B_2'\}\}$.

This proves part (b).

(c) Let $(\{j_1\}, \{k_1\})$ be a change and (B_1, B_2, B_3, B_4) the result of this change. We have

- $\min\{k \in B_3\} = j_1 < \min\{k \in B_3^0\}$;
- $\max\{j \in B_2 \text{ s.t. } j < \min\{k \in B_3\}\} = \max\{j \in B_2^0 \text{ s.t. } j < j_1\} < \max\{j \in B_2^0\} = \max\{j \in B_2^0 \text{ s.t. } j < \min\{k \in B_3^0\}\}$;

- $\max\{j \in B_2\} = k_1 > \max\{j \in B_2^0\}$;
- $\min\{k \in B_3 \text{ s.t. } k > \max\{j \in B_2\}\} = \min\{k \in B_3^0 \text{ s.t. } k > k_1\} > \min\{k \in B_3^0\} = \min\{k \in B_3^0 \text{ s.t. } k > \max\{j \in B_2^0\}\}$.

This proves that any change of one team is invalid. For changes of more than one team, use part (b) of this proposition. \square

2.1. The algorithm

With the above results in mind, we are equipped with the necessary components of an algorithm for finding all the perfect solutions with value s .

First, find all the configurations (B_1, B_2, B_3, B_4) that are valid. In order to do this, recall that B_1 is formed by the eight smallest elements of E . Then find all the possible sets B_4 by finding all the subsets of eight elements of $E \setminus B_1$ whose sum is $v - 8s$; this is carried out very easily by dynamic programming or direct enumeration.

Then, for each B_4 , find all the possible B_2, B_3 such that the configuration (B_1, B_2, B_3, B_4) is valid. First check if the basic configuration (B_1, B_2^0, B_3^0, B_4) is valid. If it is not, no changes from it can produce a valid configuration, by Proposition 2(c), so discard it and move to the next B_4 . Otherwise, if (B_1, B_2^0, B_3^0, B_4) is valid, keep it and check all the changes from it. Start with the changes of one element. By Proposition 2(a), if a change is invalid then all the changes stronger than it are also invalid. This greatly reduces the number of one-element changes to be checked. For every valid change keep the resulting configuration. Once all the changes of one element have been checked, turn to changes of two elements $(\{j_1, j_2\}, \{k_1, k_2\})$ such that all the one-element changes $(\{j_1\}, \{k_1\})$, $(\{j_1\}, \{k_2\})$, $(\{j_2\}, \{k_1\})$ and $(\{j_2\}, \{k_2\})$ are valid (by Proposition 2(b)) and analyze if they are valid, keeping the resulting configurations. Turn to three-element changes where all the two-element changes are valid and check them. Continue this procedure until all the changes of seven elements have been checked. This process is typically very fast; note, for instance, that if there are no two-element valid changes, there is no need to check any change with more than two elements by Proposition 2(b). Once all the changes from the basic configuration have been checked, move to the next B_4 . This will end with a list of all the valid configurations.

The next step is to find all the perfect solutions, if any, for each valid configuration. Recall that different configurations generate different groupings. In order to find all the perfect solutions for a configuration, $(8!)^3$ possibilities must be analyzed: every element of B_2, B_3 , and B_4 must be assigned to a group. In practice, a much smaller number of groupings is analyzed. In fact, note that groupings are generated from configurations (see Definition 4) in such a way that, for each group, the element from B_2 is smaller than the element from B_3 which is, in turn, smaller than the element from B_4 . This immediately reduces the number of possibilities to analyze. Moreover, if after the assignment of an element from B_2 (or two elements from B_2 and B_3) to a group, this group will necessarily have a score greater than or smaller than s , we need not assign elements to the rest of the groups.

Once all the perfect solutions have been found, the last step is to check if they satisfy the geographical rule to get the final list.

Table 1
Possible values of B_4 with sum of ranks equal to 224

B_4 set	Teams
B_{AA}	24, 25, 26, 27, 29, 30, 31, 32
B_{AB}	23, 25, 26, 28, 29, 30, 31, 32
B_{AC}	22, 25, 27, 28, 29, 30, 31, 32
B_{AD}	21, 26, 27, 28, 29, 30, 31, 32
B_{AE}	23, 24, 27, 28, 29, 30, 31, 32

Table 2
Teams from different geographical areas for the 2014 FIFA World Cup

Africa	Asia	Europe	North/Central America	South America
Côte d’Ivoire (17)	Japan (28)	Spain (2)	USA (13)	Brazil (1)
Ghana (22)	Iran (29)	Germany (3)	Mexico (23)	Argentina (4)
Algeria (25)	Korea Republic (30)	Belgium (6)	Costa Rica (24)	Colombia (5)
Nigeria (26)	Australia (31)	Switzerland (8)	Honduras (27)	Uruguay (7)
Cameroon (32)		Netherlands (9)		Chile (12)
		Italy (10)		Ecuador (21)
		England (11)		
		Portugal (14)		
		Greece (15)		
		Bosnia and Herzegovina (16)		
		Croatia (18)		
		Russia (19)		
		France (20)		

Australia is affiliated to the Asian Football Confederation. No team from the Oceania zone qualified. We assign Brazil rank 1, although its rank is 11, since it is the host team and, therefore, it is the seed in group A following FIFA rules.

2.2. The particular case of $E = \{1, \dots, 32\}$.

The algorithm above can be applied with any set E . If we use relative ranks, we get $E = \{1, \dots, 32\}$ and $B_1 = \{1, \dots, 8\}$. We now show the application of our algorithm to this case.

As shown above, the minimum value of a perfect solution in this case is, at least, 38. In order to find all the valid configurations with value $s = 38$, by Lemma 1, we first find all the subsets of eight elements of $\{9, \dots, 32\}$ whose sum is $32 \times 33/2 - 38 \times 8 = 224$. The only possibilities are shown in Table 1. For all these B_4 the corresponding basic configurations are valid for value 38. Moreover, for all of them, the only valid one-element changes for value 38 are $(\{16\}, \{17\})$ and $(\{16\}, \{18\})$. By Proposition 2(b), no changes of more than one element are valid for value 38. Therefore, we have 15 valid configurations for value 38; they give rise to a total of 7,629,120 perfect solutions with value 38.

For the 2014 World Cup, the teams are shown in Table 2. The numbers in parentheses are the team relative ranks. For this case, among the 7,629,120 perfect solutions, 173,160 satisfy the geographical rule. The total procedure of finding all the perfect solutions with our algorithm takes around 1 minute and 45 seconds on an AMD Athlon 64 3800+ 2.4 GHz. That is, our algorithm is able to enumerate all feasible solutions in less time than a binary programming problem model can find a single solution.

Table 3
A result of the “perfect” draw

Group
A Brazil (1), England (11), Nigeria (26), Japan (28)
B Spain (2), Côte d’Ivoire (17), Russia (19), Algeria (25)
C Germany (3), Portugal (14), Ecuador (21), Cameroon (32)
D Argentina (4), Italy (10), Costa Rica (24), Australia (31)
E Colombia (5), USA (13), France (20), Korea Republic (30)
F Belgium (6), Netherlands (9), Mexico (23), Iran (29)
G Uruguay (7), Greece (15), Bosnia and Herzegovina (16), Honduras (27)
H Switzerland (8), Chile (12), Croatia (18), Ghana (22)

Table 3 shows a result of the draw; note that in fact the sum of the ranks of the three strongest teams in each group is 38.

3. The 3-2-3 draw

The algorithm in Section 2 provides all the perfect solutions. As described in the Introduction, the practical implementation of the algorithm would require the publication of all the perfect solutions and a final draw to select one of them. This final draw would be very different to what is common in FIFA (and regional Football Associations) draws, where teams, in the form of balls, are extracted from pots. In this section, we define an alternative draw system using pots, which produces results close to the perfect solutions of Section 2. We use relative ranks from 1 to 32 throughout this section, although the same ideas can be used with arbitrary integer ranks.

Our 3-2-3 system is a modification of the S-curve-type draw system introduced by Guyon (2015). We first briefly describe the S-curve-type draw system and then explain our proposal. In the S-type system, there are four initial pots 1–4: teams ranked 1–8 go to the first pot, teams 9–16 to the second, teams 17–24 to the third, and teams 25–32 to the fourth. Each pot is then divided into two pots U and L , where the four strongest teams of pot i go to pot i_U and the four weakest teams to pot i_L . Groups A–D will be made from pots $1_U, 2_L, 3_U, 4_L$, and groups E–H from pots $1_L, 2_U, 3_L, 4_U$. In order to make the groups, balls will be drawn from pot 1_U and each team will go to one group, then balls will be drawn from pot 2_L and each team will go to one group and so on. However, this procedure would most likely produce groups that do not satisfy the geographical rule. In order to take into account the geographical constraints, there is a preliminary step before drawing the balls. This step consists of listing all the feasible continental distributions compatible with the structure of the pots, that is, the different ways in which the continents of the teams can appear satisfying the geographical rule. There will be n_U valid continental distributions for pots $1_U, 2_L, 3_U, 4_L$, and n_L valid continental distributions for pots $1_L, 2_U, 3_L, 4_U$ (in the 2014 FIFA World Cup, $n_U = 7$ and $n_L = 4$). The draw system goes as follows: first, draw a number between 1 and n_U for the continental distribution of pots $1_U, 2_L, 3_U, 4_L$, and another number between 1 and n_L for pots $1_L, 2_U, 3_L, 4_U$. Once the continental distributions are fixed, the above procedure of drawing balls is carried out: for groups A–D, balls are sequentially drawn from pot 1_U and the extracted team goes to the first group available for its continent; then balls are drawn from pot 2_L and so on. The same is done for

Table 4

Composition of the pots for the 3-2-3 draw

Group	Pot 1	Pot 2	Pot 3	Pot 4
A-C	$1_U = \{1, 2, 3\}$	$2_L = \{14, 15, 16\}$	$3_M = \{20, 21, 22\}$	$4_M = \{27, 28, 29\}$
D-E	$1_M = \{4, 5\}$	$2_U = \{9, 10\}$	$3_L = \{23, 24\}$	$4_U = \{25, 26\}$
F-H	$1_L = \{6, 7, 8\}$	$2_M = \{11, 12, 13\}$	$3_U = \{17, 18, 19\}$	$4_L = \{30, 31, 32\}$

groups E–H, with pots $1_L, 2_U, 3_L, 4_U$. For more details on the S-curve-type draw, the reader can consult Guyon (2015).

The S-curve-type draw produces very good results (much better than the current FIFA draw). However, it can still produce some undesirable results. For instance, in the 2014 FIFA World Cup, a possible outcome of this draw has a group with Spain (2), Portugal (14), Côte d’Ivoire (17), and Iran (29) scoring 33 (with the sum of the ranks of the three strongest teams) and another group with Switzerland (8), Chile (12), Costa Rica (24) and Japan (28), scoring 44. That means that a score difference of 11 between two groups is possible with this draw system. Our 3-2-3 draw will avoid these differences among groups. We now describe our 3-2-3 draw system.

As in the S-curve-type system, we introduce the eight strongest teams into pot 1, teams from 9 to 16 into pot 2, teams 17–24 into pot 3, and teams 25–32 into pot 4. The difference is how we divide the pots. Each pot is divided into three pots (instead of two) in the following way: $1_U = \{1, 2, 3\}$, $1_M = \{4, 5\}$, $1_L = \{6, 7, 8\}$, $2_U = \{9, 10\}$, $2_M = \{11, 12, 13\}$, $2_L = \{14, 15, 16\}$, $3_U = \{17, 18, 19\}$, $3_M = \{20, 21, 22\}$, $3_L = \{23, 24\}$, $4_U = \{25, 26\}$, $4_M = \{27, 28, 29\}$, $4_L = \{30, 31, 32\}$. Then groups A–C will be drawn from pots $1_U, 2_L, 3_M, 4_M$, groups D–E will be drawn from pots $1_M, 2_U, 3_L, 4_U$, and groups F–H will be drawn from pots $1_L, 2_M, 3_U, 4_L$. Table 4 shows the pots for this draw.

Draws with these pots produce very balanced solutions, with very similar group scores. However, the results of the draw may not satisfy the geographical rule. In order to get solutions satisfying this rule we first find the feasible continental distributions. Like in Guyon (2015), we find all the feasible continental distributions for the three parts of the draw: there will be n_U, n_M, n_L of them, respectively. It might be the case, as in Guyon’s system, that some n_i are 0; that is, there are no continental distributions compatible with the composition of the pots. For instance, if seven teams in pots $1_U, 2_L, 3_M, 4_M$ are from Europe, they cannot be put into three groups with no more than two European teams in each group. We then say that the composition of the pots is infeasible. Another problem (like in Guyon’s scheme) occurs when two teams are predetermined to play in the same group, regardless of the result of the draw. This is regarded as undesirable because some groups are partially fixed beforehand. In that case, we also consider the composition of the pots as infeasible. When the composition of the pots is infeasible, we propose to swap some teams in the pots. For instance, we swap teams 29 and 30 in pot 4. That means that team 29 will be in pot 4_L and team 30 will be in pot 4_M . If the resulting composition is still infeasible, we return teams 29 and 30 to their original pots and swap teams 26 and 27 and check if the resulting composition is feasible. That is, we keep swapping teams until the resulting composition is feasible. There are many ways to carry out this procedure; we describe an algorithm in the Appendix.

Once a feasible composition of the pots is found, all the continental distributions for each of the three draws are listed and the day of the draw one is picked at random for each draw. Then balls

Table 5
Composition of the pots for the 3-2-3 draw

Group	Pot 1	Pot 2	Pot 3	Pot 4
	Pot 1 _U	Pot 2 _L	Pot 3 _M	Pot 4 _M
A-C	1 Brazil 2 Spain 3 Germany	13 USA 15 Greece 16 Bosnia and Herzegovina	20 France 21 Ecuador 22 Ghana	27 Honduras 28 Japan 29 Iran
	Pot 1 _M	Pot 2 _U	Pot 3 _L	Pot 4 _U
D-E	4 Argentina 5 Colombia	9 Netherlands 10 Italy	23 Mexico 24 Costa Rica	25 Algeria 26 Nigeria
	Pot 1 _L	Pot 2 _M	Pot 3 _U	Pot 4 _L
F-H	6 Belgium 7 Uruguay 8 Switzerland	11 England 12 Chile 14 Portugal	17 Côte d'Ivoire 18 Croatia 19 Russia	30 Korea Republic 31 Australia 32 Cameroon

are sequentially extracted and teams go to the first available group according to the continental distribution.

In the particular case of the 2014 FIFA World Cup, the original composition in Table 4 is infeasible because there are two European teams in pot 1_U and three European teams in pot 2_L; therefore, Brazil and France are predetermined to be in the same group. The algorithm in the Appendix first tries to swap 29–30, which also produces an infeasible composition, since the infeasibility is due to pot 2. The algorithm continues and the first feasible composition is obtained by swapping teams 13 (USA) and 14 (Portugal). The composition is shown in Table 5.

The number of feasible continental distributions for this composition are $n_U = 16$, $n_M = 1$, $n_L = 4$. An example of feasible continental distribution is given in Table 6 and a result of the draw with this continental distribution is given in Table 7; note that the sum of the three strongest teams in each group ranges from 36 to 40.

Remark 3. *Adding more randomness to the draw.* Our 3-2-3 draw guarantees that no teams are predetermined to play together. However, the eight teams in pots 1_M, 2_U, 3_L, and 4_U will be forming groups D–E, so there are not many different results for these groups. A possibility to increase the randomness of this draw is to define several compositions of pots by slightly changing these pots: swap at most one team from each of the pots 1_M, 2_U, 3_L, and 4_U with a team from 1_L (or 1_U), 2_M, 3_M, 4_M, respectively, and pick one of the compositions at random before the final draw.

4. Comparison of the draws

We have developed two draw systems with the objective of making groups with very similar strength. In this section, we compare our systems with two other existing systems: the FIFA draw system and the S-curve-type draw system proposed in Guyon (2015).

Table 6

A feasible continental distribution for the composition in Table 5.

Group	Pot 1	Pot 2	Pot 3	Pot 4
	Pot 1 _U	Pot 2 _L	Pot 3 _M	Pot 4 _M
A	Brazil	Europe	Europe	Asia
B	Spain	Europe	Africa	North/Central America
C	Germany	North/Central America	South America	Asia
	Pot 1 _M	Pot 2 _U	Pot 3 _L	Pot 4 _U
D	Argentina	Europe	North/Central America	Africa
E	Colombia	Europe	North/Central America	Africa
	Pot 1 _L	Pot 2 _M	Pot 3 _U	Pot 4 _L
F	Belgium	South America	Europe	Asia
G	Uruguay	Europe	Europe	Africa
H	Switzerland	Europe	Africa	Asia

Table 7

A result of the 3-2-3 draw

Group	Teams	Score
A	Brazil (1), Greece (15), France (20), Japan (28)	36
B	Spain (2), Bosnia and Herzegovina (16), Ghana (22), Honduras (27)	40
C	Germany (3), USA (13), Ecuador (21), Iran (29)	37
D	Argentina (4) Netherlands (9), Costa Rica (24), Algeria (25)	37
E	Colombia (5), Italy (10), Mexico (23), Nigeria (26)	38
F	Belgium (6), Chile (12), Croatia (18), Australia (31)	36
G	Uruguay (7), England (11), Russia (19), Cameroon (32)	37
H	Switzerland (8), Portugal (14), Côte d'Ivoire (17), Korea Republic (30)	39

In order to make the comparison, we analyze the possible results of the different draw systems for the 2014 World Cup. For each result we compute the score of each group, defined as the sum of its three strongest teams. Thus, for each result we have eight scores (one per group). The more similar the eight scores, the better the result, because the groups are more balanced. Therefore, for each result we compute its range (the greatest group score minus the smallest group score) and the standard deviation of the eight group scores.

For the perfect draw system in Section 2, the eight group scores are equal to 38, so all the results have range and standard deviation equal to zero. For the FIFA, S-curve-type and 3-2-3 draw systems we have simulated 10,000 results. We summarize the outcomes in Table 8. Also Fig. 1 shows (a smoothed version of) the histograms of the standard deviations for the three draw systems. Of course, the perfect draw system of Section 2 is the best, since both the range and standard deviation of each result are 0. When comparing the other three systems, we observe that the FIFA draw system is much worse than the other two and that the 3-2-3 outperforms the S-curve-type. In particular, observe that in the S-curve-type system, a result may have a difference equal to 11 between the score of the strongest group and the score of the weakest group, while this maximum difference is reduced to 7 in the 3-2-3 system. The maximum difference in the FIFA draw system is 37.

Table 8

Location measures for the range and standard deviation (SD) of group scores

	Range				SD			
	Min	Mean	Median	Max	Min	Mean	Median	Max
FIFA	6	21.61	22	37	1.73	6.99	7.03	11.53
S-curve-type	5	8.52	8	11	2.06	2.77	2.78	3.57
3-2-3	1	4.75	5	7	0.50	1.56	1.58	2.45

The score of a group is the sum of the ranks of its three strongest teams.

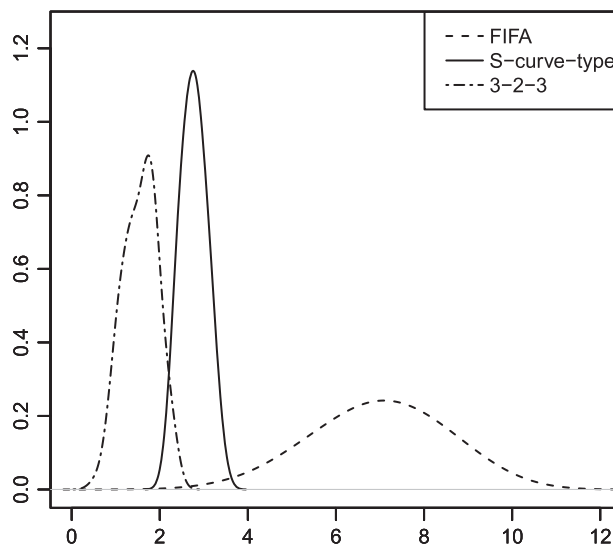


Fig. 1. Smoothed histograms of the standard deviation of group scores. The score of a group is the sum of the ranks of its three strongest teams.

One question that may arise is how the perfect and the 3-2-3 systems would perform if, instead of defining the score of a group as the sum of the ranks of its three strongest teams, we defined it as the sum of the ranks of its four teams. It is not clear that our systems work well in that situation, because they have been designed for the definition of the three strongest teams. Nevertheless, we compare them with the FIFA and S-curve-type draws, using that definition. In this case, for the perfect draw system, the scores of the eight groups will no longer be equal, so we also compute the range and standard deviation of the results of that system. Since we know all the possible outcomes of the perfect draw system (the 173,160 perfect solutions in Section 2) we do not need to simulate it and we directly compute the range and standard deviation for each result. For the rest of the systems, we simulate 10,000 results. The data are summarized in Table 9 and Fig. 2. We do not represent the histogram of the standard deviation in the perfect draw system because it takes only five different values (2.74, 2.91, 3, 3.08, 3.24).

Again our draw systems perform much better than the FIFA system. The S-curve-type performs slightly better than our proposals. As explained above, this happens because our draw systems are designed to get balanced results when the score of the group is defined by its three strongest teams.

Table 9

Location measures for the range and standard deviation (SD) of group scores

	Range				SD			
	Min	Mean	Median	Max	Min	Mean	Median	Max
FIFA	6	28.78	29	54	2.12	9.07	9.11	15.72
S-curve-type	1	5.79	6	11	0.50	2.15	2.12	3.84
Perfect	8	9.45	9	11	2.74	3.01	3.00	3.24
3-2-3	6	8.13	8	11	2.12	2.68	2.69	3.61

The score of a group is the sum of the ranks of its four teams.

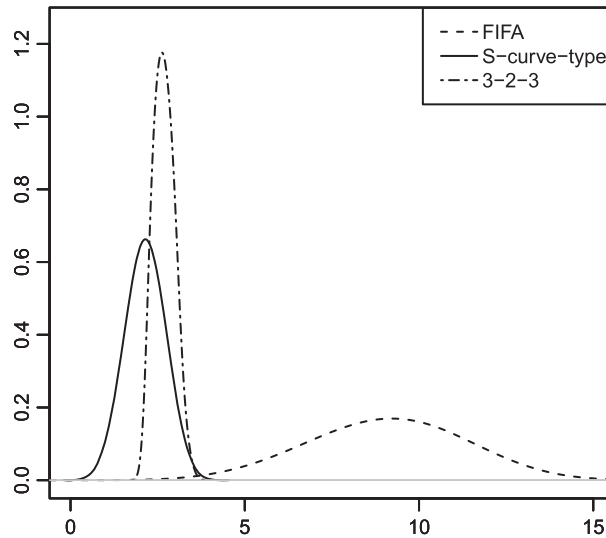


Fig. 2. Smoothed histograms of the standard deviation of group scores. The score of a group is the sum of the ranks of its four teams.

Nevertheless, note that the maximum difference between the strongest and weakest group is 11 in the three draw systems (S-curve-type, perfect, and 3-2-3). In Fig. 2 we observe that the histograms of the standard deviations for the S-curve-type and the 3-2-3 are much closer to each other than in Fig. 1. Thus, in a global comparison between the 3-2-3 and the S-curve-type draw systems, we prefer the 3-2-3, since the gain when the score of a group is defined in terms of its three strongest teams is greater than the loss when considering its four teams.

5. Conclusions

We have introduced two draw systems that produce results where the groups have very similar (or equal) scores, defined as the sum of the ranks of their three strongest teams. This definition of score of a group is a measure of its strength (the lower the score, the stronger the group) and it is

considered as more relevant than the sum of the ranks of all the teams in the group; see Section 2 in Guyon (2015).

Our systems can be used for any tournament with seeding and geographical rules where 32 teams are split into eight groups (FIFA World Cup, UEFA Champions League, etc.). They can also be used for sports other than soccer. Slight modifications of the systems would be needed for a different number of teams or groups. One of the systems, called the perfect draw system, produces results where all the groups have exactly the same strength. The other system, called the 3-2-3 draw system, produces results where the groups have very similar strength. Moreover, this system is better suited for a TV show than the perfect draw system.

Our two systems work much better than the system used by FIFA. They also outperform the S-curve-type draw system proposed in Guyon (2015). Even though our systems are designed to provide good results when the definition of strength of a group is based on the ranks of its three strongest teams, they also work well if the weakest team is included in the definition. With this definition, our systems still work much better than the FIFA draw system and only slightly worse than the S-curve-type draw system in Guyon (2015).

As a conclusion, we recommend the use of the perfect draw system, which produces perfectly balanced groups, or, if a traditional draw with pots is preferred, the 3-2-3 draw, which produces results very close to perfect balance.

Acknowledgments

We thank the Associate Editor and the referees for their useful comments and suggestions. F. Javier López acknowledges financial support from projects MTM2014-53340-P and MTM2017-83812-P of MINECO (Spain). He is a member of the research group Modelos Estocásticos (DGA).

References

- Aisch, G., Leonhardt, D., 2014. Mexico, the World Cup's luckiest country. *The New York Times*. Available at <http://www.nytimes.com/2014/06/06/upshot/mexicos-run-of-world-cup-luck-has-continued.html> (last accessed January 15, 2018).
- Alarcón, F., Durán, G., Guajardo, M., 2014. Referee assignment in the Chilean football league using integer programming and patterns. *International Transactions in Operational Research* 21, 3, 415–438.
- Durán, G., Guajardo, M., Sauré, D., 2017. Scheduling the South American qualifiers to the 2018 FIFA World Cup by integer programming. *European Journal of Operational Research* 262, 1109–1115.
- Guyon, J., 2014. La FIFA doit aussi revoir le tirage au sort de sa Coupe du monde. *Le Monde*. Available at http://www.lemonde.fr/coupe-du-monde/article/2014/06/04/il-faut-repenser-le-tirage-au-sort-de-la-coupe-du-monde_4431571_1616627.html (last accessed January 15, 2018).
- Guyon, J., 2015. Rethinking the FIFA World CupTM final draw. *Journal of Quantitative Analysis in Sports* 11, 169–182.
- Jones, M.C., 1990. The World Cup draw's flaws. *The Mathematical Gazette* 74, 335–338.
- Kantar Media, 2015. 2014 FIFA World Cup BrazilTM. Television Audience Report. Available at [https://resources.fifa.com/mm/document/affederation/tv/02/74/55/57/2014fwcbrazilvaudiencereport\(draft5\)\(issuedate14.12.15\)_neutral.pdf](https://resources.fifa.com/mm/document/affederation/tv/02/74/55/57/2014fwcbrazilvaudiencereport(draft5)(issuedate14.12.15)_neutral.pdf) (last accessed January 15, 2018).
- Korf, R.E., 2009. Multi-way number partitioning. Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence, IJCAI09, pp. 538–543.
- Lasek, J., Szlávik, Z., Bhulai, S., 2013. The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition* 1, 1, 27–46.

- Law, M., Bagchi, R., 2014. Champions league group-stage draw 2017: Tottenham to play Real Madrid, Chelsea face Atletico, Manchester United in with Benfica. *The Telegraph*. Available at <http://www.telegraph.co.uk/football/2017/08/24/champions-league-group-stage-draw-2017-time-tv-channel-teams/> (last accessed January 15, 2018).
- McHale, I., Davies, S., 2007. Statistical analysis of the effectiveness of the FIFA world rankings. *Statistical Thinking in Sports*, Vol. 15. Chapman & Hall/CRC, Boca Raton, FL, pp. 77–90.
- McMahon, B., 2013. Why the FIFA 2014 World Cup Finals will be unique and very unfair. *Forbes Magazine*. Available at <https://www.forbes.com/sites/bobbymcmahon/2013/12/01/why-the-fifa-2014-world-cup-finals-will-be-unique-and-very-unfair/#78b2c4842dab> (last accessed January 15, 2018).
- Menezes, J., 2017. Rugby World Cup draw: England drawn in “Pool of Death” against France and Argentina at Japan 2019. *Independent*. Available at <http://www.independent.co.uk/sport/rugby/rugby-union/international/rugby-world-cup-draw-2019-england-pool-of-death-france-argentina-a7727521.html> (last accessed January 15, 2018).
- Michiels, W., Korst, J., Aarts, E., Van Leeuwen, J., 2003. Performance ratios for the differencing method applied to the balanced number partitioning problem. *Lecture Notes in Computer Science* 2607, 583–595.
- Prodinger, H., 1982. Number of partitions of $(1, \dots, n)$ into two sets of equal cardinalities and equal sums. *The Canadian Mathematical Bulletin* 25, 2, 238–241.
- Prodinger, H., 1984. On the number of partitions of $\{1, \dots, n\}$ into r sets of equal cardinalities and sums. *Tamkang Journal of Mathematics* 15, 161–164.
- Ribeiro, C.C., 2012. Sports scheduling: problems and applications. *International Transactions in Operational Research* 19, 1–2, 201–226.
- Wikipedia. 2017. Group of death. Wikipedia. Available at https://en.wikipedia.org/wiki/Group_of_death (accessed December 17, 2017).
- Woeginger, G.J., Yu, Z., 1992. On the equal-subset-sum problem. *Information Processing Letters* 42, 6, 299–302.
- Zhang, J., Mouratidis, K., Pang, H., 2011. Heuristic algorithms for balanced multi-way number partitioning. *IJCAI Proceedings—International Joint Conference on Artificial Intelligence*, Vol. 22, pp. 693–698.

Appendix

As explained in Section 3, the initial composition of the pots shown in Table 4 may not be feasible as no outcome of the draw would produce a result satisfying the geographical rule. We also consider a composition as infeasible if there are two teams that are predetermined to be in the same group, irrespective of the result of the draw. This infeasibility problem was also present in the S-curve-type system proposed by Guyon (2015); see Section 8 in that paper.

We describe an algorithm which, starting from the composition of Table 4, swaps teams within pots 1–4 between pots L, M, U in order to find a feasible composition. The algorithm allows for a maximum of one swap of two teams within each pot 1–4. Therefore, we define a change as an 8-tuple

$$(i_1, i_2, j_1, j_2, k_1, k_2, l_1, l_2),$$

with $1 \leq i_1 \leq i_2 \leq 8$, $9 \leq j_1 \leq j_2 \leq 16$, $17 \leq k_1 \leq k_2 \leq 24$, $25 \leq l_1 \leq l_2 \leq 32$. The meaning of a change is that, from the initial composition in Table 4, teams i_1 and i_2 are swapped, teams j_1 and j_2 are swapped, teams k_1 and k_2 are swapped, and teams l_1 and l_2 are swapped. Note that we allow, for example, $i_1 = i_2$, which means that no teams are swapped in pot 1. For instance, the change $(3,3,12,16,23,23,27,32)$ produces the composition shown in Table A1.

The algorithm is very simple. It keeps trying different changes from the composition in Table 4 until a feasible composition is found. The changes are tried in an order designed to preserve, as

Table A1

Composition of the pots for the 3-2-3 draw after the change (3,3,12,16,23,23,27,32)

Group	Pot 1	Pot 2	Pot 3	Pot 4
A-C	$1_U = \{1, 2, 3\}$	$2_L = \{12, 14, 15\}$	$3_M = \{20, 21, 22\}$	$4_M = \{28, 29, 32\}$
D-E	$1_M = \{4, 5\}$	$2_U = \{9, 10\}$	$3_L = \{23, 24\}$	$4_U = \{25, 26\}$
F-H	$1_L = \{6, 7, 8\}$	$2_M = \{11, 13, 16\}$	$3_U = \{17, 18, 19\}$	$4_L = \{27, 30, 31\}$

much as possible, the good properties of the 3-2-3 draw system. For a change, we define the vector of characteristics

$$(d, c_7, c_6, c_5, c_4, c_3, c_2, c_1, i_2 - i_1, j_2 - j_1, k_2 - k_1, l_2 - l_1, -i_2, -j_2, -k_2, -l_2),$$

where $d = i_2 - i_1 + j_2 - j_1 + k_2 - k_1 + l_2 - l_1$, $c_r = \mathbf{1}_{\{i_2 - i_1 = r\}} + \dots + \mathbf{1}_{\{l_2 - l_1 = r\}}$, for $r = 1, \dots, 7$ (here $\mathbf{1}$ stands for the indicator function). Note that there is a one-to-one correspondence between the changes and the vector of characteristics (no two different changes have the same vector of characteristics). We use the lexicographical order for the vector of characteristics; that is, the algorithm will try the changes in increasing lexicographical order of their vector of characteristics until it finds a feasible composition. The rationale for this lexicographical order is that we want a composition similar to that in Table 4, since draws from that composition produce very well balanced solutions. In that spirit, we prefer changes with a small d , which can be seen as a measure of discrepancy between the composition in Table 4 and the resulting composition. Within the same discrepancy, we prefer more “evenly distributed changes”; for instance, for a discrepancy of $d = 4$, we prefer two swaps of teams whose difference is 2 (13–15 and 22–24, say) to a swap of two teams whose difference is 4 (25–29). Within this, we prefer changes in pots with a higher number (a swap 26–30 in pot 4 is preferred to a swap 10–14 in pot 2). Last, within this, we prefer changes in a pot to be made in teams with higher numbers (swap 26–30 is preferred to swap 25–29).

It might be the case that the algorithm tries all the changes without finding a feasible composition of the pots. That is, no change defined by swapping at most two teams in each pot 1–4 yields a feasible composition. In that case, a more elaborate algorithm, allowing for instance more than one swap in each pot or swapping teams between pots 1–4 would be necessary. We do not consider this situation here because it is very unlikely; we have tested our algorithm for the teams (and ranks) of the FIFA World Cups in 1998, 2002, 2006, 2010, and 2014, and in all the cases a feasible composition of the pots was found in the first steps of the algorithm.